Data processing steps for DQH_blood_CuffLinks_Genome_data.xls:

The Illumina BCL output files were converted to fastq-sanger file format and sequence quality trimming was performed using Trimmomatic on iPlant Collaborative's Discovery Environment using the High-Performance Computing applications. The following Trimmomatic parameters were used: ILLUMINACLIP:TruSeq3-SE.fa 2:30:10; LEADING:10; TRAILING:10; SLIDINGWINDOW:4:20; HEADCROP:6; MINLEN:36. Reads were mapped to the Ensembl Tursiops truncatus genome, turTru1 v76.1, using Tophat2 v 2.3.13 with bowtie2 v 2.2.4 as the alignment engine on iPlant Collaborative's Discovery Environment using the High-Performance Computing applications. Read counts as FPKM (fragments per kilobase of transcript per million mapped reads) were generated using Cufflinks v 2.2.0 with the genome as a reference on iPlant Collaborative's Discovery Environment using the High-Performance Computing applications.

Processed data is supplied in a single file. The first column "Gene ID" contains the gene ID from the Ensembl genome. The second column "Gene Symbol" contains the gene symbol, if assigned, from the Ensembl turTru1 v76.1 annotation. The remaining columns contain the FPKM values for each sample as generated by Cufflinks. A FPKM > 0 in at least half the samples and average FPKM ≥ 1 across all samples was required for all further data analysis. Only genes meeting these requirements are included in the processed data table.

Data processing steps for DQH_blood_RSEM_Genome_data.xls:

The Illumina BCL output files were converted to fastq-sanger file format and sequence quality trimming was performed using Trimmomatic on iPlant Collaborative's Discovery Environment using the High-Performance Computing applications. The following Trimmomatic parameters were used: ILLUMINACLIP:TruSeq3-SE.fa 2:30:10; LEADING:10; TRAILING:10; SLIDINGWINDOW:4:20; HEADCROP:6; MINLEN:36. Reads were mapped to the Ensembl *Tursiops truncatus* genome, turTru1 v76.1, using RSEM v 1.2.18 with bowtie2 as the alignment engine and read counts were generated as FPKM (fragments per kilobase of transcript per million mapped reads) at the gene level.

Processed data is supplied in a single file. The first column "Gene ID" contains the gene ID from the Ensembl genome. The second column "Gene Symbol" contains the gene symbol, if assigned, from the Ensembl turTru1 v76.1 annotation. The remaining columns contain the FPKM values for each sample as generated by RSEM. A FPKM > 0 in at least half the samples and an average FPKM ≥ 1 across all samples was required for all further data analysis. Only genes meeting these requirements are included in the processed data table.

Data processing steps for DQH_blood_RSEM_Trinity_data.xls:

The Illumina BCL output files were converted to fastq-sanger file format and sequence quality trimming was performed using Trimmomatic on iPlant Collaborative's Discovery Environment using the High-Performance Computing applications. The following Trimmomatic parameters were used: ILLUMINACLIP:TruSeq3-SE.fa 2:30:10; LEADING:10; TRAILING:10; SLIDINGWINDOW:4:20; HEADCROP:6; MINLEN:36. The read files from one summer and one winter globin-depleted sample from each animal (n=8; Hua: Feb and Sept, Kainalu: Feb and Aug, Keo: Feb and Aug, Pele: Feb and Sept) were concatenated into a single fastq file for assembly using a minimum K-mer coverage of 1, a minimum overlap value of 25 and a minimum contig length of 400 nucleotides on iPlant Collaborative's Discovery Environment using the High-Performance Computing applications. Annotation of the de novo assembly

was obtained by BLASTx searches of the human subset of the uniprot_swissprot database. Reads were mapped to the de novo Trinity assembly, using RSEM v 1.2.18 with bowtie2 as the alignment engine and read counts were generated as FPKM (fragments per kilobase of transcript per million mapped reads) at the gene level.

Processed data is supplied in a single file. The first column "Gene ID" contains the gene ID from the *de novo* Trinity assembly. The second column "Accession Number" contains the uniprot_swissprot accession number, if assigned, from BLASTx searches. The remaining columns contain the FPKM values for each sample as generated by RSEM. A FPKM > 0 in at least half the samples and an average FPKM ≥ 1 across all samples was required for all further data analysis. Only genes meeting these requirements are included in the processed data table.

Data processing steps for Female_v_Male_FC.xlsx:

FPKM values generated by RSEM were used to calculate fold change values, at the gene level, in EBSeq between female (n=16, all samples from Keo and Pele) and male (n=16, all samples from Hua and Kainalu) samples. The first column “Gene ID” contains the gene ID from the *de novo* Trinity assembly. The second column “Log_2 Fold Change” contains the Log_2(fold change) value calculated by EBSeq. Positive Log_2(fold change) values are more highly expressed in females. The third column “PPDE” is the EBSeq calculated posterior probability that a gene is differentially expressed. With the hard-threshold method used to control FDR, only those genes with PPDE ≥ 1 – FDR are significantly differentially expressed.

Data processing steps for Sum_v_Win_FC.xlsx:

FPKM values generated by RSEM were used to calculate fold change values, at the gene level, in EBSeq between summer (n=8, Hua: July and August, Kainalu: July and September, Keo: August, Pele: July, August, and September) and winter (n=8, Hua: December and February, Kainalu: December and February, Keo: December and February, Pele: December and February) samples. The first column “Gene ID” contains the gene ID from the *de novo* Trinity assembly. The second column “Log_2 Fold Change” contains the Log_2(fold change) value calculated by EBSeq. Positive Log_2(fold change) values are more highly expressed in warmer months. The third column “PPDE” is the EBSeq calculated posterior probability that a gene is differentially expressed. With the hard-threshold method used to control FDR, only those genes with PPDE ≥ 1 – FDR are significantly differentially expressed.

Data processing steps for Sum_v_Win_Male_FC.xlsx:

FPKM values generated by RSEM were used to calculate fold change values, at the gene level, in EBSeq between summer (n=4, Hua: July and August, Kainalu: July and September) and winter (n=4, Hua: December and February, Kainalu: December and February) in samples from male animals. The first column “Gene ID” contains the gene ID from the *de novo* Trinity assembly. The second column “Log_2 Fold Change” contains the Log_2(fold change) value calculated by EBSeq. Positive Log_2(fold change) values are more highly expressed in warmer months. The third column “PPDE” is the EBSeq calculated posterior probability that a gene is differentially expressed. With the hard-threshold method used to control FDR, only those genes with PPDE ≥ 1 – FDR are significantly differentially expressed.

Data processing steps for Sum_v_Win_Female_FC.xlsx:

FPKM values generated by RSEM were used to calculate fold change values, at the gene level, in EBSeq between summer (n=4, Keo: August, Pele: July, August, and September) and winter (n=4, Keo: December and February, Pele: December and February) samples. The first column "Gene ID" contains the gene ID from the *de novo* Trinity assembly. The second column "Log_2 Fold Change" contains the Log_2(fold change) value calculated by EBSeq. Positive Log_2(fold change) values are more highly expressed in warmer months. The third column "PPDE" is the EBSeq calculated posterior probability that a gene is differentially expressed. With the hard-threshold method used to control FDR, only those genes with PPDE ≥ 1 – FDR are significantly differentially expressed.